Political Analysis (2025), **33**, 393–411
doi:10.1017/pan.2025.5

PA

ARTICLE

Seeing Like a District: Understanding What Close-Election Designs for Leader Characteristics Can and Cannot Tell Us

Andrew Bertoli[1] and Chad Hazlett[2]

[1]School of Politics, Economics, and Global Affairs, IE University, Segovia, Castile and León, Spain; [2]Departments of Political Science and Statistics, University of California, Los Angeles, CA, USA

Corresponding author: Chad Hazlett; Email: chazlett@ucla.edu

(Received 18 December 2023; revised 1 October 2024; accepted 22 November 2024; published online 27 March 2025)

Abstract

Many influential political science articles use close elections to study how important outcomes vary after a certain type of candidate wins, such as a Democrat or a Republican. This politician characteristic regression discontinuity (PCRD) design offers opportunities for inferential leverage but also the potential for confusion. In this article, we clarify what causal claims the PCRD licenses, offering a rigorous causal analysis that points to three principal lessons. First, PCRDs do nothing to isolate the effect of the politician characteristic of interest as apart from other politician characteristics. Second, selection processes (regarding both "who runs" and "which elections are close") can generate and exacerbate such confounding, as noted in Marshall (2024). Third and more fortunately, this approach does make it possible to estimate the average effect of *electing* a leader of type "A" vs. "B" in the context of close elections, treating the units as districts, not leaders. We also suggest a set of tools that can aid in falsifying key assumptions, avoiding unwarranted claims, and surfacing mechanisms of interest. We illustrate these issues and tools through a reanalysis of an influential study about what happens when extremists win primaries (Hall 2015).

Keywords: regression discontinuity design; close-election design; causal inference; potential outcomes

Edited by: Dr. Jeff Gill

1. Introduction

In the last two decades, many studies have examined how important political and economic outcomes differ following close elections between different types of candidates, referred to as the "politician characteristic regression discontinuity" (PCRD) design by Marshall (2024). As illustrated in Table 1, one large group of these studies analyzes the relationship between the political party of leaders and various outcomes, such as taxation (Fredriksson, Wang, and Warren 2013), unemployment (Leigh 2008), and racial inequality (Beland 2015). A second commonly studied leader attribute has been gender, with a number of studies examining how electing women influences various health, educational, and political outcomes (Bhalotra, Clots-Figueras, and Iyer 2018; Brollo and Troiano 2016; Broockman 2014; Clots-Figueras 2012; Ferreira and Gyourko 2014; Jankowski, Marcinkiewicz, and Gwiazda 2019). Other studies use this approach to study a wider range of attributes, including religion (Bhalotra *et al.* 2014; Hopkins and McCabe 2012), political extremism (Hall 2015), tribal connections (Xu and Yao 2015), and business background (Szakonyi 2021).

Joint first authors

Table 1. Selected PCRD studies

Study	Attribute	Outcome(s)	Context
Leigh (2008)	Political Party	Unemployment	United States
Gerber and Hopkins (2011)	Political Party	Spending Allocation	United States
Fredriksson *et al.* (2013)	Political Party	Taxation	United States
Beland (2015)	Political Party	Racial Inequality	United States
Erikson, Folke, and Snyder Jr. (2015)	Political Party	Election Outcome	United States
Benedictis-Kessner and Warshaw (2016)	Political Party	Total Spending	United States
Broockman and Ryan (2016)	Political Party	Citizen Engagement	United States
Thompson (2020)	Political Party	Immigration Enforcement	United States
Dynes and Holbein (2020)	Political Party	Economic Outcomes	United States
Benedictis-Kessner and Warshaw (2020)	Political Party	Total Spending	United States
Thomas (2018)	Political Party	Pork Distribution	North India
Callen, Gulzar, and Rezaee (2020)	Political Party	Public Services	Pakistan
Brollo and Nannicini (2012)	Political Party	Money Transfers	Brazil
Novaes (2018)	Political Party	Party Fragility	Brazil
Feierherd (2020)	Political Party	Presidential Election Outcome	Brazil
Bertoli, Dafoe, and Trager (2019)	Political Party	Military Conflict	International
Ferreira and Gyourko (2014)	Gender	Various Political Outcomes	United States
Broockman (2014)	Gender	Women's Political Participation	United States
Clots-Figueras (2012)	Gender	Education	India
Bhalotra and Clots-Figueras (2014)	Gender	Child Mortality	India
Bhalotra *et al.* (2018)	Gender	Future Female Candidates	India
Brollo and Troiano (2016)	Gender	Corruption	Brazil
Hopkins and McCabe (2012)	Religion	Police Hiring	United States
Bhalotra *et al.* (2014)	Religion	Health & Educational Outcomes	India
Hall (2015)	Political Extremism	General Election Outcome	United States
Xu and Yao (2015)	Tribal Connections	Public Investment	China
Szakonyi (2021)	Business Background	Spending Allocation	Russia

Despite the seemingly straightforward nature of this prevalent research approach, confusion persists over what it is able to reveal and how to correctly interpret its results. In this article, we clarify what precisely is identified and estimated by PCRDs, the correct and incorrect interpretation of their results, and what scientific and policy value they provide. In short, we argue against two views.

First, as we show below, investigators have often described PCRD results as estimates of the causal effect of politician attributes, such as the "effect of gender" or the "effect of ideology" (e.g., Huidobro and Falcó-Gimeno 2023, 1563; Núñez and Dinas 2023, 1; Carozzi and Gago 2023, 380). This way of thinking about PCRD stems from viewing the units of analysis as the elected leaders (what we call the "leader-as-unit" approach) and imagining their counterfactual outcomes had they been the other type. In the first part of our analysis, we formally describe the causal estimand that results from this approach and show that PCRDs clearly do not identify it. In so doing, we also formalize what PCRDs are equipped

to identify: the local average treatment effect (LATE) of districts *electing* a candidate with the attribute in question, which is clearly identified by PCRDs (under what we call the "district-as-unit" approach). While our analysis provides a clear and novel explanation of why PCRDs cannot estimate the causal effect of politician attributes, the point is not new. Researchers have been making it since at least Sekhon and Titiunik (2012), and it is reinforced in the recent contribution of Marshall (2024). Further, while some researchers have claimed that PCRDs identify the causal effects of politician attributes, others have been careful to avoid such language and to correctly interpret the results as we advocate below (e.g., Broockman 2014; Hall 2015, 2019).

The second issue we address stems from how researchers interpret the result given that it does not represent the effect of the attribute of interest itself. The PCRD can identify the "effect of electing a candidate of type A vs. B," which combines the effect of differences between winners of the two groups on a surfeit of related variables, further complicated by the selection process that put pairs of candidates into close elections (Marshall 2024). Marshall (2024, 508–509) argues that the apparent complexity of this quantity limits its value to researchers and thus the value of PCRD designs. We argue to the contrary. This apparent complexity is a symptom of comparing what PCRDs can identify to a target quantity that should not in fact be the target. It is not necessary or even desirable to untangle the effect of the candidate characteristic from these other variables, and the presumed effect of the attribute is itself difficult or impossible to define in many settings. Rather, in many cases the combined effect that PCRDs do identify is precisely the question of scientific or policy relevance, because it speaks to what can be potentially manipulated in the future and what can be expected given the (narrow) election of a given type of candidate. Hall (2019, 30–33) makes a similar point in the context of studying what happens when extremists win primaries. We extend this discussion to apply to PCRDs in general.

Along with clarifying the confusion currently surrounding PCRD designs, we contribute to the PCRD literature in several other ways. First, researchers have previously recognized that PCRDs should balance district-level covariates (e.g., Patterson Jr. 2020, 4–5; Marshall 2024, 506). This is correct and relates to the identifiability of the effect of *districts (barely) electing* a candidate of a given type rather than another. We extend this point, noting that PCRD will similarly balance any characteristic not affected/selected by the election outcomes—including candidate characteristics. For example, PCRDs will balance pre-election candidate-level factors for both candidates, such as the total campaign money spent by type A politicians or whether the type A bare winners and losers were incumbents. This implies a larger number of possible balance tests that can be used to falsify the close-election design or to surface chance imbalances in the observed sample. To our knowledge, Hall (2015) is the only study in the PCRD literature to examine balance on such pre-election candidate-level factors. Second, we point out that some PCRD designs categorize politicians as type A or type B by using a non-binary variable, like political ideology scores or age, to sort candidates into two groups. We call such approaches "gradient PCRDs" and consider some issues and opportunities they raise. Third, we outline a set of tools and diagnostics that can be used when analyzing PCRD designs, some of which apply to any regression discontinuity setting, and some which apply specifically to PCRDs.

Our article proceeds as follows. In Section 2, we explain why PCRD designs identify the effect of electing certain types of politicians but not the effect of politician attributes. We also explain why estimating the effect of electing certain types of politicians is valuable from a theoretical and policy perspective. In Section 3, we describe a number of tools for analyzing PCRDs, including falsification tests that the close-election design is working as expected and descriptive analyses that illustrate confounding of the "treatment" characteristic with other politician characteristics to aid in correct interpretation. In Section 4, we illustrate both the lessons of our analysis and these tools, examining the impact of nominating extremist candidates for U.S. house races on who wins the general election (Hall 2015).

2. Understanding the PCRD Design

We begin by imagining a large number of close elections between type A and type B candidates. The distinction might be that type A candidates graduated from college and type B candidates did not. We

will first conceptualize the treatment as the type A candidate winning. For district i, we will use the indicator variable $D_i \in \{0,1\}$ to denote whether the type A candidate won. Similarly, we will denote district i's potential outcome if the type A candidate won by $Y_i(1)$ and its potential outcome if the type B candidate won by $Y_i(0)$. The individual-level treatment effect for district i is then

$$\tau_i = Y_i(1) - Y_i(0), \tag{1}$$

and we are interested in the average of such district-level effects, local to districts at the cut-point ($mov = 0$),

$$\bar{\tau}_{districts} = \mathbb{E}\left[Y_i(1) - Y_i(0) | mov = 0\right]. \tag{2}$$

Note that i indexes districts in the close-election group, and this expectation is taken over those districts. This quantity contemplates "what would have happened if each district had elected its type A candidate (compared to if it had elected its type B candidate)." We refer to this quantity and what it represents as the "district-as-unit" view.

Contrast with the "Effect of the Attribute"

We contrast this to a different target quantity that investigators have pursued, and which is responsible for many complications and much confusion in this literature. For this estimand, the politicians (winners) are the units indexed by i (rather than districts), and the potential outcomes are $Y_i(A)$ (the outcome for politician i had they been of type A) and $Y_i(B)$ (the outcome for politician i had they been of type B). That is, the causal effect contrasting these potential outcomes compares an individual politician's outcome had that politician been of type A vs. type B. This approach presupposes that elected politicians have well-defined potential outcomes had they been the other type, which may not be a realistic assumption in many contexts (see below). However, we will assume momentarily that counterfactuals are well-defined, just to write the estimand that results. The LATE would then be,

$$\bar{\tau}_{leaders} = \mathbb{E}\left[Y_i(A) - Y_i(B) | mov = 0\right]. \tag{3}$$

This quantity can also be labeled as arising from the "leader-as-unit" view, since i indexes elected politicians and contemplates a contrast of potential outcomes (by type) within politician.

Figure 1 offers a visual illustration of these potential outcomes and contrasts between them. We continue with the running example of "graduating from college" as the characteristic of interest. Let us suppose for example that whether candidates graduated from college had a small impact on the outcome, while other differences between the candidates who did and did not graduate from college had a large impact on the outcome. The left panel illustrates the "district-as-unit" view, defining an effect of districts electing a candidate of each type, $\bar{\tau}_{districts}$. The right panel of Figure 1 visualizes the "leader-as-unit" approach, showing $\bar{\tau}_{leaders}$. In this case, $\bar{\tau}_{districts}$ is much larger than $\bar{\tau}_{leaders}$, which would be the average of $\mathbb{E}\left[Y_i(A) - Y_i(B) | mov = 0_-\right]$ and $\mathbb{E}\left[Y_i(A) - Y_i(B) | mov = 0_+\right]$.[1]

Use of These Estimands in the PCRD Literature

One challenge with the PCRD literature, and any effort to describe it, is that it contains a diversity of interpretations and ambiguities in language. Neither view regarding the above estimands is universally practiced. Numerous important studies have understood the PCRD approach to estimate a local average

[1] We use "$mov = 0_-$" and "$mov = 0_+$" to indicate that these quantities are limits taken from below ($mov < 0$) or above ($mov > 0$). That is, for example,

$$\mathbb{E}\left[Y_i(A) | mov = 0_+\right] \equiv \lim_{\epsilon \downarrow 0} \mathbb{E}\left[Y_i(A) | mov = \epsilon\right],$$

$$\mathbb{E}\left[Y_i(A) | mov = 0_-\right] \equiv \lim_{\epsilon \uparrow 0} \mathbb{E}\left[Y_i(A) | mov = \epsilon\right].$$

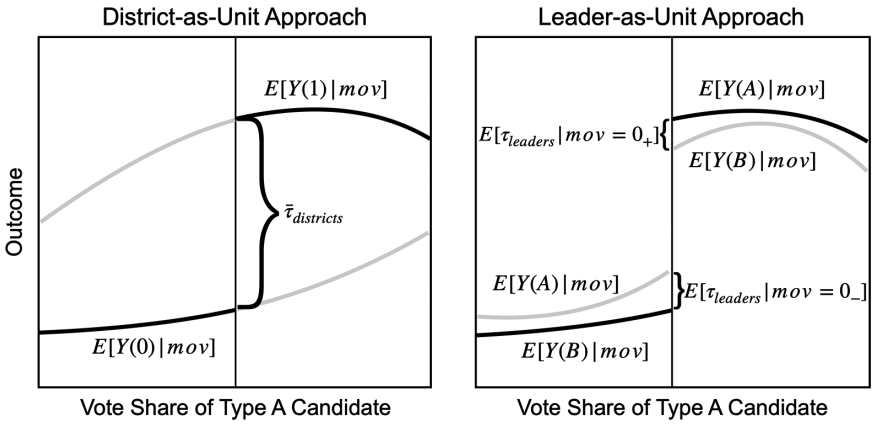

Figure 1. Graphical illustrations of the potential outcomes and estimands. In the left-hand panel, districts are the units, and the contrast at *mov* = 0 gives the local average treatment effect for districts electing their type A candidates (instead of their type B candidates). In the right-hand panel, the politician is the unit, and the contrasts ($\tau_{leaders}$) consider the effect of the elected politician "being type A" (compared to "being type B") at *mov* = 0. Thus, in this panel we assume that elected politicians have well-defined counterfactual outcomes had they been the other type, a notion that is likely to be problematic in many PCRD contexts. If such counterfactuals within politicians are well-defined, $\bar{\tau}_{leaders}$ would be the average of $\mathbb{E}[\tau_{leaders}|mov = 0_-]$ and $\mathbb{E}[\tau_{leaders}|mov = 0_+]$.

treatment effect of districts electing a candidate of type A vs. type B, including well-known and widely cited works such as Broockman (2014), Dynes and Holbein (2020), and Hall (2015). From this perspective, the notion that it could be understood otherwise may even seem perplexing.

Yet, one can also readily find examples of PCRD studies that describe their results as if they are estimating the effect of a politician having the attribute in question ($\bar{\tau}_{leaders}$), or analyses that describe methodological challenges as though the effect of the attribute was the intended target. For example, numerous articles comparing close elections between women and men claim to estimate the "effect of gender" (e.g., Casarico, Lattanzio, and Profeta 2022, 9; Huidobro and Falcó-Gimeno 2023, 1563; Carozzi and Gago 2023, 380). Similarly, studies examining close elections between candidates from different parties often claim to estimate the "effect of ideology" (e.g., Alonso and Andrews 2020, 748; Núñez and Dinas 2023, 1).

More systematically, we reviewed all the PCRD studies that we could identify published in *APSR*, *AJPS*, and *JOP* since 2020, which include a total of 28 articles. We found that 17 of the 28 (61%) make inconsistent statements about the parameter that PCRD identifies, claiming or suggesting at times that it identifies the causal effect of an attribute and at other times that it identifies the effect of electing a type A (instead of a type B) candidate. We provide a list of quotes from these studies in the online appendix. In some cases, this may be a problem of ambiguous use of language rather than clearly referencing $\bar{\tau}_{leaders}$ as opposed to $\bar{\tau}_{districts}$. This is no less important to address, as it leaves the reader uncertain or possibly mistaken about what is being estimated. Further, confusion on this matter may have increased given recent debates. In critiquing this literature, Marshall (2024, 501) proceeds under the premise that the goal of PCRD designs is "to isolate effects of a specific characteristic of elected politicians on downstream outcomes." As we argue, many of the concerns raised therein regarding the complications of interpreting PCRD can be seen as symptoms of first regarding $\bar{\tau}_{leaders}$ as the target.

2.1. Which of These Estimands Should Investigators Be Concerned with?

Our purpose in defining these two quantities is not to suggest that the researcher has the choice between them, but to organize different views that have generated confusion and sometimes incorrect interpretations in the literature. We next argue that the effect of electing a leader of one type vs. the

other ($\bar{\tau}_{districts}$) should be the target of investigation, whereas the effect of the attribute within politicians ($\bar{\tau}_{leaders}$) is neither what the PCRD identifies, nor is it necessarily desirable to learn, or even well-defined.

Bias and Identification

Identification of $\bar{\tau}_{districts}$ is straightforward: the revelation mechanism provided by the running variable (*mov*) allows us to see (or approximate) the average outcome among districts electing a politician of type A at $mov = 0$ ($\mathbb{E}[Y(1)|mov = 0]$) and that for districts electing a politician of type B ($\mathbb{E}[Y(0)|mov = 0]$). With both of these identifiable in the data, $\bar{\tau}_{districts} = \mathbb{E}[Y(1) - Y(0)|mov = 0]$ is thus identifiable. The formal identification requirement, thus, is simply an assurance that $\mathbb{E}[Y(1)|mov = 0]$ can be unbiasedly estimated using a model built on data "from the right," $\hat{\mathbb{E}}[Y(1)|mov = 0_+]$, and likewise $\mathbb{E}[Y(0)|mov = 0]$ can be unbiasedly estimated using a model from the left, $\hat{\mathbb{E}}[Y(0)|mov = 0_-]$. These two conditional expectations are represented by the two dark lines in the left panel of Figure 1, and the above assumption can be posed as a smoothness assumption for both of these lines as they approach $mov = 0$ (Imbens and Lemieux 2008, 618).

By contrast, nothing about the nature of close elections provides us with the necessary inferential leverage to understand what would have happened had a politician of either type been of the opposite type. In Figure 1, we can see that the conditional expectation functions of the politicians' potential outcomes ($\mathbb{E}[Y(A)|mov]$ and $\mathbb{E}[Y(B)|mov]$) are discontinuous at $mov = 0$. Thus, the closeness of an election offers no leverage for estimating an effect of politicians switching between types. If an investigator was hoping to identify the effect of the attribute ($\bar{\tau}_{leaders}$), then even under a suitable continuity assumption on $\mathbb{E}[Y_i(0)|mov = 0_-]$ and $\mathbb{E}[Y_i(1)|mov = 0_+]$, the bias separating this from its estimate under PCRD is

$$\mathbb{E}[\hat{\tau}_{leaders} - \bar{\tau}_{leaders}] = \tfrac{1}{2}\bar{\tau}_{districts} + \tfrac{1}{2}\left(\mathbb{E}[Y_i(A)|mov = 0_-] - \mathbb{E}[Y_i(B)|mov = 0_+]\right). \quad (4)$$

We refer readers to the online appendix for the derivation. This bias term could point in either direction and be large in magnitude. In Figure 1, our estimate for the causal effect of the attribute would be about five times greater than its actual effect.

Ontological Challenges to $\bar{\tau}_{leaders}$

As alluded to above, proposing to estimate the effect of politician attributes ($\bar{\tau}_{leaders}$)—or interpreting one's PCRD design as if it does so—can pose deeper challenges than the bias and identification concerns. Specifically, this estimand may not reflect a meaningful or interesting quantity as the counterfactuals it invokes may be nonsensical. For instance, what would it mean for a male politician to be a female politician or vice versa? At what point in a politician's history are we imagining their counterfactual sex, and in what sense is it "changed"? Similarly, what do we mean by the effect of being a Democrat vs. Republican? Is it the underlying beliefs and ideology, the network personal commitments, their donor bases, their status in the majority or minority party, or something else? There is a further question of who becomes a politician. In our running example, if a given politician who graduated college would not have run for office had they not graduated, then contemplating $Y_i(B)$ for that person is problematic as they would not have held office if they had been type B. Defining a potential outcome for this person would require some kind of manipulation that keeps them in office despite changing this characteristic. Such a counterfactual notion may be defined in some philosophical frameworks but is clearly awkward and does not correspond to a simple and clear hypothetical experimental manipulation or "ideal experiment."

Further, even if these concerns can be set aside, we argue that the effect of the attribute ($\bar{\tau}_{leaders}$) is rarely the quantity of interest to investigators, whereas the effect of electing a politician of that type ($\bar{\tau}_{districts}$) is often of value. For example, suppose researchers, lobbyists, or citizens wish to know what to expect regarding debt spending or tax policy under a (narrowly) elected Republican. They likely wish to know how the outcome will differ from what is expected if their Democratic opponent had instead been

elected. This is the counterfactual of relevance. Even if it were possible or well-defined, we are likely less interested in a notion of what would happen if the Republican winner could instead somehow be made into a Democrat, without altering that politician's educational background, committee positions, etc. We thus disagree with the argument that the quantity identifiable by PCRD is often too complicated or uninteresting to study (Marshall 2024).

Heterogeneous Bundles and SUTVA

Two concerns that may arise in interpreting $\bar{\tau}_{districts}$ are bundling and heterogeneity. To be clear, candidate attributes in PCRD designs like party and gender are highly "bundled" categories that serve as summary labels for some largely unnamed set of underlying correlated characteristics (Marshall 2024, 508). However, the solution to this is found in desisting from regarding the effect of the attribute ($\bar{\tau}_{leaders}$) as the object of first interest. The bundling is in fact "desirable" in the settings described thus far, because a pure manipulation of only "party" or "gender" is of unclear interest at best. As noted, if we wish to know what would have happened on average, had the Republican rather than the Democrat barely been elected, we do not mean that we narrowly change their party without also seeing the concomitant changes on other characteristics like ideology. Focusing on $\bar{\tau}_{districts}$ and the "district-as-unit" view helps to keep this firmly in mind.

A second related concern is that of variation in what treatment means. No two districts elect the same type A or type B candidate. In fact, there can be substantial variation between candidates of a given type across districts. As such, PCRDs may at first be thought to violate the uniform treatment component of the Stable Unit Treatment Value Assumption (SUTVA) (Marshall 2024, 508). Such issues, however, have nothing to do with PCRD in particular; they arise commonly in the social sciences. For example, several famous studies use the biological sex of newborn babies as a treatment (e.g., Glynn and Sen 2015; Dube and Harish 2020), like to test how having a younger sister (instead of a younger brother) affects a person's political attitudes later in life (Healy and Malhotra 2013). Clearly, no one has the same younger sister (or younger brother), except people who are in the same family. Similarly, experiments on intergroup contact theory typically compare subjects who were randomized to interact with one of two types of people (e.g., Broockman and Kalla 2016; Mousa 2020). The distinction might be gender, race, religion, or sexual orientation. Regardless, there are many differences across the type A people in such studies, as well as across the type B people, and it can be difficult to define exactly what it means to be type A vs. type B. The nature of the intergroup contact—whether it goes well or not, for example—is also variable.

The treatment in PCRD designs—electing the type A candidate—is no different in this regard and does not pose a unique obstacle to accumulating scientific knowledge. Rather, in each of these cases the careful consideration of SUTVA assumptions lead us not to claim that inference is impossible, but to be disciplined in how we understand the treatment. Here, among districts with close elections, either the candidate of type A or the candidate of type B could have won and both of those outcomes are well-defined for unit i, whatever those characteristics might mean for unit i. The consistency assumption holds: $Y_i(1)$ is the outcome we would observe for district i had the type A candidate won, and $Y_i(0)$ would be the observed outcome had the candidate of type B won. The contrast of these two (potential) outcomes is thus well-defined as well, even if "type" would have meant something different in a different unit. We refer readers to more thorough treatments of this concern such as in VanderWeele and Hernan (2013).

Why Compensating Differentials Do Not Threaten Internal Validity

Our analogy to intergroup contact theory also clarifies another of Marshall's critiques of PCRD designs, which centers on "compensating differentials." Marshall (2024) explains that type A and type B candidates in close races may differ from type A and type B candidates in less competitive races. For example, if the electorate is biased against women, then women in close races against men should (on average) be more competent than their male opponents (Marshall 2024, 497). Like a PCRD design,

an experiment on intergroup contact theory randomizes subjects to interact with type A or type B individuals who may be atypical, because they are people who agreed to help with the study. Therefore, selection processes determining which type A and type B individuals assist with the study could create, amplify, or reduce average differences between type A and type B individuals compared to the overall population. Once we understand the estimand to be the effect of electing politicians of a given type—not the effect of the type on the politician—this ceases to threaten internal validity. It remains, however, of critical importance for considering causal mechanisms and external validity, which Marshall (2024) rightly raises attention to. In this sense we agree with his suggestion that the issue might be viewed as a limitation to the "external validity and interpretability of PCRD estimates" (Marshall 2024, 508), rather than the claims elsewhere (e.g., 498, 499, 504, and 509) that it threatens internal validity.

Beyond compensating differentials, we also note that it could make little sense to extrapolate too far away from the cut-point (for example, imagining a very unpopular candidate getting elected, or a Republican getting elected in an 80% Democratic district). Such counterfactuals are implausible, but also of limited scientific or policy interest in many cases, since a decision between supporting one candidate type or another would not arise in such cases (see Hall 2015). Both of these reasons serve as reminders for researchers to carefully attend to the local nature of the PCRD estimand and estimate.

2.2. The Value of Studying the Effects of Electing Certain Types of Politicians

We now return to the claim that the effect of electing certain types of politicians, rather than the effect of the attribute itself, is in fact of direct interest in many contexts.

Identification of Politician Effects

PCRDs most obviously shed light on what we might call "agency" or "politician" effects. For example, Gerber and Hopkins (2011) examine close mayoral elections between Democrats and Republicans from 1990 to 2006. Their estimates suggest that who wins has a large impact on certain types of government spending, including roads, housing, and public safety (Gerber and Hopkins 2011, 335), highlighting the important impact that mayors can exert on city budgets. Likewise, Bertoli, Dafoe, and Trager (2024) find that countries are much less likely to engage in military conflicts after barely electing older rather than younger national leaders. This finding highlights that leaders have an important impact on foreign policy despite the structural constraints imposed on states by the international system (Waltz 2010). In both of these examples, neither compensating differentials nor the bundled nature of treatment prevents these PCRD designs from establishing that leaders clearly matter. Rather, such issues are crucial for thinking about the mechanisms that explain the observed politician effects and considering whether such effects might be different in less-competitive races.

By allowing researchers to identify politician effects (or their absence in certain contexts), PCRDs can also shed light on other important theoretical questions that go beyond simply whether and how politicians matter. We will now turn to several examples.

Median Voter Theorem

PCRD designs have helped quantify the failure of the median voter theorem to explain U.S. politics. Studies have shown that the roll-call voting of districts tends to differ greatly depending on whether Republicans or Democrats win close elections (Fowler and Hall 2017; Lee, Moretti, and Butler 2002). In particular, Fowler and Hall (2017, 358) find that, on average, the outcomes of these narrow races change the probability that the districts' representatives will vote conservative by about 40 percentage points. In short, representatives from these districts tend not to converge to the median voter, instead largely sticking to their divergent ideological positions. The theoretical significance of this finding is not negated by compensating differentials or the complex nature of what it means to be a Democrat or a Republican.

Retrospective Voting

Dynes and Holbein (2020) examine whether important social and economic outcomes differ when Democrats or Republicans control state governments, including after close races between Democratic or Republican gubernatorial candidates (A85–A103). They find surprisingly little difference in their outcome variables 2–4 years after each election, suggesting that the effects of party control will likely not be evident to voters before the subsequent elections are held. These results may imply that voters cannot effectively hold parties accountable, or alternatively that under theoretically perfect accountability the parties may converge in policy (Dynes and Holbein 2020). Regardless of interpretation, the causal effect of interest here is that of party control (e.g., whether the Democratic or Republican gubernatorial candidate wins), not the causal effect of ideology per se. If ideology did have a large impact on the outcomes within 2-4 years, but its effect was largely canceled out by other differences between Democratic and Republican governors, then voters would still be unable to engage in reliable retrospective voting based on the key social and economic indicators that Dynes and Holbein (2020) examine.

Political Polarization

Hall (2019) uses PCRD to support his theory that the rise in political polarization in U.S. Congress since 1980 is largely driven by a decline in the number of moderate candidates wanting to run for office. In particular, Hall (2019, 15–26) argues that fewer moderates have run in recent decades because of the rising costs and shrinking benefits of holding office, not because voters prefer extremist candidates and therefore deter moderates from running. To rule out the latter possibility, Hall (2019, 52–56) examines close primary races between moderate and extremist candidates from 1980 to 2012. He finds that moderate bare winners were far more likely than extremist bare winners to prevail in the subsequent general election. Thus, voters prefer moderate candidates when given the choice, suggesting that the rise in polarization in the U.S. House has much more to do with who wants to run for office than voter preferences. Importantly, the causal effect of extremism per se is not nearly as relevant to Hall's theory (Hall 2019, 4). The important question with respect to his theory is whether voters deter moderate candidates from running, because of ideological preferences or for other reasons (Hall 2019, 4–5).

The Policy Importance of the Estimand

From a policy perspective, PCRD designs also provide valuable insights. They can shed light on what to expect when certain types of candidates win close elections over other types of candidates. For example, Gerber and Hopkins (2011) point to how public spending could differ under Democratic and Republican mayors, while Dynes and Holbein (2020) caution against thinking that party control of state governments will have a large impact on key economic and social indicators in the short or medium-term. PCRD designs can also provide parties with insights about which candidates to nominate and voters with insights about who to elect. For instance, Hall (2019) draws attention to the potential risk for parties of nominating extremist candidates, who have historically been much more likely to lose the general election compared to moderate candidates. Similarly, research showing the benefits of electing women on health outcomes (e.g., Bhalotra and Clots-Figueras 2014) might help voters decide who to support in a close race between a man and woman.

2.3. Final Observations About PCRD Designs

Before moving on to outline diagnostic tools that researchers can use when analyzing PCRDs, we will discuss two important points that are largely unrecognized in the PCRD literature.

What Is Comparable in Close Elections?

It has been well-noted that PCRD designs balance "district-level" covariates (e.g., Patterson Jr. 2020, 4–5; Marshall 2024, 506). By balance on some X in this context, we mean there would be no discontinuity

found in expectation for (the mean of) that X at the cutoff. This clarifies that, excepting chance imbalances, the effect we estimate with PCRD cannot be due to differences between the kinds of districts that (barely) elect a leader of type A vs. type B. Indeed, this feature helps to address a major concern that would arise in comparisons that do not rely on close elections. Further, district characteristics are just an example of variables that are not affected (or selected) by the election outcome. All such features can be expected to be similar between elections barely won by candidates of type A or B and will thus be balanced (as noted in the case of standard RDs by Imbens and Lemieux 2008). Such variables also include, for example, all candidate-level characteristics that do not depend on the election outcome (e.g., "whether the female candidate was the incumbent" in a design comparing male to female leaders). Not included in this set would be any variable that does depend on the outcome, such as "whether the male vs. female winner was the incumbent." We note that in about half of districts near the cut-point, the value of this type of variable will refer to a pre-election characteristic of a candidate who was not actually the winner, and for this reason such data are not always collected. However, where these variables can be collected, they add to the set of balance tests investigators can run as falsification tests or to be alert to finite sample imbalances.

Gradient PCRDs

Some PCRD studies examine attributes that are initially non-binary. Analysis then requires dichotomizing these into binary variables, sometimes using one global rule to determine "high" vs. "low" values, and other times using within-district comparisons. Further, many studies employ a "reverse caliper," meaning that they only include a district in the analysis if the candidates running there differ by at least a certain amount on the (original, continuous) attribute. For instance, Bertoli *et al.* (2024) dichotomize age into "older" or "younger" based on whether candidates were older or younger than their main electoral opponent, and subset to the elections where the candidates were at least 10 years apart in age. Similarly, Hall (2015) defines candidates as extremists or moderates relative to the candidate who they ran against in their primary election, subsetting to cases where the ideological gap between the two primary candidates was at or above the median for all competitive primary races.

Such designs, which we call "gradient PCRDs," do not pose additional problems for statistical inference under the "district-as-unit" approach to PCRD, as the parameter of interest is still well-defined and identified. However, several issues are noteworthy. First, the potential for a non-monotonic relationship between the initial non-binary variable and the outcome should be carefully considered when interpreting the results and could make researchers more likely to get false negatives. Second, in cases where the dichotomization is determined by a within-district comparison, this must be remembered during interpretation. For example, a candidate labeled as "younger" in one district might be older than the "old" candidate in another. The third issue is actually an opportunity: where a reverse caliper is applied, it means that only a subset of data is used for the main effect estimation. However, as we show below, many of the additional diagnostics and falsification tests we describe can be applied to the larger set of close elections, providing greater power to detect potential problems.

3. Tools and Diagnostics

We describe here statistical tools that are useful either for falsifying PCRD's identifying assumptions, or for shedding interpretive light on the estimate. We organize these in three categories: tools applicable to any RD design, those specific to PCRDs, and additional tests to note regarding gradient PCRDs.

3.1. Useful Tools for Any RD

Many of the common tools used to analyze RDs can be applied to PCRDs, provided the analyst understands the district to be the unit of analysis. First, the now standard McCrary density test (McCrary 2008) can aid in detecting fine-sorting, examining whether there was an unusually high number of

districts where candidates of type A (or type B) barely won. To run this test, each district is taken as a single observation, coding whether the candidate of type A won (1) or lost (0). A jump (or drop) in the local density of districts electing type A candidates near $mov = 0$ would be cause to suspect some form of non-comparability of the districts at the threshold, as there may be some reason why type A candidates are more likely to narrowly win (or lose) these close-elections.

Second, researchers can compare districts where type A candidates barely won to districts where type B candidates barely won. In expectation, these two groups of districts should be very similar in terms of pre-election environmental factors like total population, GDP per capita, land area, prior political characteristics, etc. Let $\boldsymbol{E}$ denote such a vector of environmental covariates. Imbalances in $\boldsymbol{E}$ between the two groups of districts on these factors can signal non-randomness/ fine-sorting of potential concern (Bertoli *et al.* 2019; Hall 2015; Marshall 2024; Sekhon and Titiunik 2012).

Third, a conventional balance-testing approach would make these comparisons one variable at a time, each time using an RD estimation strategy but with the chosen variable as the "outcome," to check for a discontinuity in its expected value at $mov = 0$. This is useful, but it will miss cases where several variables in $\boldsymbol{E}$ may have small imbalances that only become statistically detectable when combined, linearly or otherwise. Further, it is possible to have balance in the means of separate variables at the threshold, without having balance on non-linear functions of those variables that can matter to the outcome and thus generate bias (see e.g., Hazlett 2020). It is even possible that multiple imbalances with opposing impacts can exist without generating bias and while improving efficiency (see Wainstein 2022). To incorporate these possibilities and avoid the limitations of such univariate balance tests, we note two useful ways that investigators could check balance on potentially worrying functions of covariates:

- *Imbalance/discontinuity in predicted treatment*. Construct an estimate of $\hat{D}_i \approx Pr[D_i|\boldsymbol{E}_i]$, the predicted probability of a candidate of type A winning ($D_i = 1$), given the vector of baseline covariates, $\boldsymbol{E}_i$. This $\hat{D}_i$ can then be used as the "outcome" in an RD analysis checking for a discontinuity in its value where *mov* crosses zero. This approach is similar to the approach proposed by Gagnon-Bartsch and Shem-Tov (2019).
- *Imbalance/discontinuity in the predicted potential outcomes*. Recall the potential outcomes defined above, with $Y_i(1)$ being the outcome district i would have if it elects a leader of type A, and $Y_i(0)$ the outcome it would have if it elects a leader of type B. Model $\widehat{Y_i(0)} \approx \mathbb{E}[Y_i(0)|\boldsymbol{E}_i]$, i.e., predict the outcome (Y_i) as a function of environmental characteristics, using only observations below the cutoff. Use this model to obtain predicted values of $\widehat{Y_i(0)}$ for all districts, on both sides of the cutoff. Then treat this variable as the outcome in an RD, checking for any discontinuity in the modeled value of $Y_i(0)$ at *mov*. The same exercise can then be repeated for $Y_i(1)$, i.e., training a model on observations above the cutoff.

In any of these cases, the model for D_i, $Y_i(1)$, or $Y_i(0)$ can be constructed using any modeling approach, so long as over-fitting is avoided or mitigated. These two tests are both useful, to different purposes. The first focuses on the probability of the type A candidate winning at the cut-off, so as to *surface evidence that the type A or type B winners in the close elections might have sorted at the cut-off*. This can be especially useful for exploration and qualitatively assessing the claim of smoothness in potential outcomes at the cutoff. However, it is not well-linked to the potential for bias in the estimate itself, because not all imbalances/sorting would lead to bias: an $\boldsymbol{E}$ that has no impact on Y can be found to be discontinuous in its expectation at $mov = 0$, without it implying any bias in the ultimate estimate with outcome Y. By contrast, the second test—checking imbalances on the *predicted potential outcomes*—is directly tuned to detecting imbalances/discontinuities in functions of $\boldsymbol{E}$ that could lead to bias.

In the context of PCRDs, a discontinuity in predicted $Y_i(1)$ or $Y_i(0)$ is straightforward to interpret. Imagine a study that looked at unemployment levels after Democrats vs. Republicans barely won close elections, where the forcing variable was the Democrat's margin of victory. A jump in the modeled expectation of $Y_i(1)$ at the cut-point would suggest that in the types of districts where Democrats barely won close elections, the (predicted) level of unemployment under a Democrat was higher than in the types of districts where Democrats barely lost close elections. A jump in $Y_i(0)$ would suggest the exact

opposite, but for Republicans. Therefore, in terms of bias concern, it is most troubling when the two imbalances go in the same direction (meaning both go up, or both go down discontinuously at the cut-point). More generally, the bias in the RD estimator of the LATE is an average of these two sources of bias,

$$\begin{aligned}\mathbb{E}[\hat{\tau}_{RD} - \bar{\tau}_{RD}] &= Pr(D=1)Bias(LATT) + Pr(D=0)Bias(LATC) \\ &= Pr(D=1)[\mathbb{E}[Y(1)|mov=0_+] - \mathbb{E}[Y(1)|mov=0_-]] + \\ &\quad Pr(D=0)[\mathbb{E}[Y(0)|mov=0_+] - \mathbb{E}[Y(0)|mov=0_-]],\end{aligned} \tag{5}$$

where LATT is the local average treatment effect among the treated and LATC is the local average treatment effect among the controls. The resulting bias in the RD estimator would be small if the imbalance in both $Y_i(1)$ and $Y_i(0)$ are small, or if they are of similar magnitudes but in opposite directions.

3.2. Tools Specific to the PCRD

We now describe two tools relevant to the specific design of the PCRD. Consider first a comparison of the characteristics of bare winners of type A to those of type B. For the "district-as-unit" estimand described here, differences (imbalances) on these characteristics among bare winners do not invalidate the design. Rather they *suggest possible explanations for why electing a candidate of type A might lead to different outcomes than electing a candidate of type B.* Performing such analyses will not only aid in proposing reasons for the observed difference, but also vividly remind readers that this approach compares counterfactuals within set districts, not counterfactuals within set leaders. For example, if type A is women and type B is men, then we might see notable differences between the male and female candidates on party, ideology, whether the candidate was the incumbent, and many other factors.

Second, balance tests can be performed to examine the differences on characteristics between bare winners and losers, separately for those of type A and those of type B. Beyond potentially casting doubt on the presumed randomness of close elections, these tests could reveal chance differences between bare winners and bare losers of a given type that could shed light on the results. For instance, suppose a much higher percentage of the female candidates who barely won had children compared to the female candidates who barely lost. Even if we believe this discrepancy is merely a chance difference, it might be informative in explaining the results. On a practical note, such balance checks require collecting information on winning and losing candidates. While this process might be time consuming, incumbency status stands out as a key variable that should usually be easy to obtain for each candidate. An imbalance on this factor could raise important concerns about sorting.

Currently, very few studies collect information on losing candidates (see Bertoli *et al.* 2019 and Bertoli *et al.* 2024 for rare exceptions). Hall (2015), which we use as our example below, does this in part, providing information on the extremists in each primary election, but not information on the moderates.

3.3. Additional Tests for Gradient PCRDs

Statistical falsification tests are themselves vulnerable to limitations of statistical power, so opportunities to improve statistical power in these tests are valuable. In gradient PCRD, while the elections included in the primary analysis may be few due to the nature of the comparison, the wider dataset can be employed in many of these falsification tests. For example, Bertoli *et al.* (2024) focus on elections where the top-two candidates were at least 10 years apart in age. While it is certainly necessary to run falsification tests on this sub-sample, many of these tools can also be applied across the entire sample of close elections (including those with lesser age differences). Any important sorting by treatment type that occurs is then more likely to be detected.

We also recommend examining how results change after adjusting the subsetting threshold (Hall 2015). For instance, Bertoli *et al.* (2024) plot their main estimates when the 10-year age difference requirement is set anywhere between 1 and 20 years. It is reasonable to expect that the estimated effect will grow larger as the minimum distance between the two candidates increases, while the confidence intervals will also likely widen because the sample size will shrink as more cases are dropped from the sample (Bertoli *et al.* 2024; Hall 2015).

4. Application: "What Happens When Extremists Win Primaries?" (Hall 2015)

We now illustrate the interpretations of PCRD and the tools described above by application to Hall's influential study comparing cases where extremist candidates barely won or barely lost to moderate candidates in U.S. House primary elections from 1980 to 2010 (Hall 2015). We provide here an abbreviated analysis, leaving a fuller analysis to the online appendix. Note that our analysis largely follows Hall's original analysis, but it also extends it in places with some of the tests that we introduced in the previous section.

Hall's important study provides a clear example of a gradient PCRD. The candidate trait of interest was initially defined on a continuous scale (a measure of each candidate's left-right ideological position). Hall then dichotomizes this measure, labeling the top-two candidates in each primary election as the "more extreme" or "more moderate" one based on their estimated left-right ideological positions. The author then subsets to the cases where the absolute difference in the estimated ideological scores of the two candidates falls at or above the median.

Hall studies several outcomes, including (i) vote share in the general election, (ii) winning the general election, (iii) the district's DW-NOMINATE score in the following congressional term, and (iv) downstream outcomes from future elections. For brevity, we focus here only on whether the extremist/moderate won the general election right after the primary in question. Following Hall's analysis, we estimate the difference at the cut-point using third-order polynomial regression lines on both sides of the cut-point that span the range of the forcing variable. This differs from that of Hall (2015) in that we add the interaction terms so that the shape of the curve can differ on the left and right sides of the cut-point. In the online appendix, we also show the results using the `rdrobust` software in `R` (Calonico, Cattaneo, and Titiunik 2014; Calonico, Cattaneo, and Titiunik 2015), which produces similar findings.

Figure 2 illustrates the main results. The shaded regions represent the 95% bootstrapped confidence intervals. As the figure shows, parties were much less likely to win the general election when the extremist barely won the primary compared to when the moderate barely won. The results are statistically significant at the 5% level ($p \approx 0.011$). The estimate at the cut-point suggests that nominating the extremist candidate in the primary led to about a 45 percentage point drop in the party's likelihood of winning the general election (se ≈ 18 percentage points).

Figure 3 illustrates the results from the McCrary density test, with the horizontal axis showing how close the extremist was to winning the primary election. For the truncated sample where the top-two candidates were at or above the median in ideological distance, the distribution is apparently smooth through the cut-point. There is no evidence of sorting in either direction, suggesting that extremist candidates were not more or less likely to win close elections than moderate candidates ($p \approx 0.96$). For the full sample, the data is not quite as smooth through the cut-point, but the difference is not statistically significant ($p \approx 0.44$, two-tailed).

Next, we examine balance on environmental factors, which is anticipated to be good if the continuity of potential outcomes assumption holds. The dark blue point estimates and confidence intervals in Figure 4 show the balance for a number of covariates originally included in Hall's dataset, along with region, which we added. In the truncated dataset, the balance on these factors looks good, with the only concerning finding being that extremists were less likely to barely win in the Northeast ($p \approx 0.004$, two-tailed). This level of imbalance is about what we would expect by chance given multiple comparisons. Using instead the full sample, none of the differences are statistically significant at the 5% level, despite

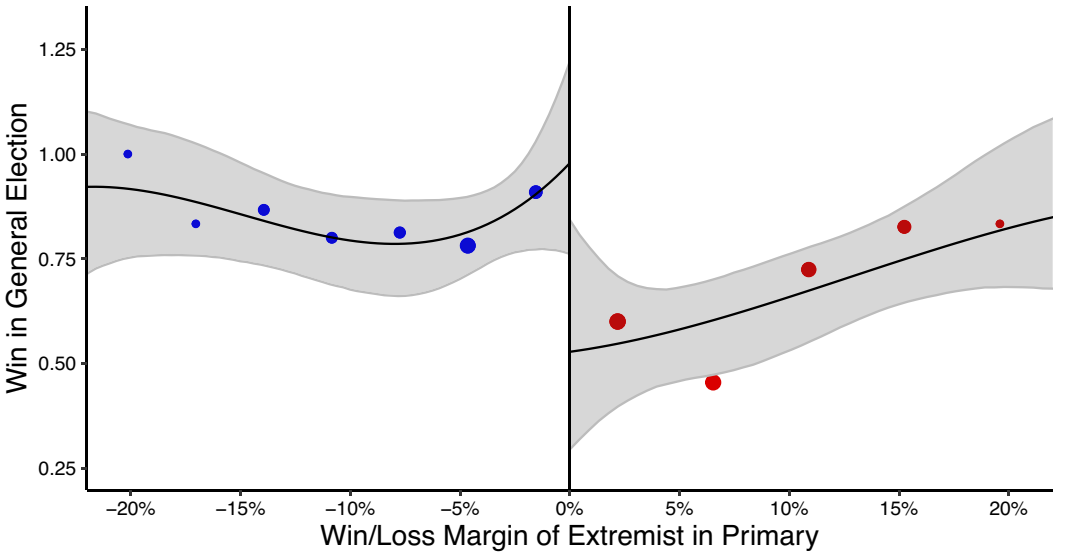

Figure 2. Regression discontinuity graph illustrating the impact of nominating the extremist candidate in the primary on the party's likelihood of winning the general election (1980–2010, $n = 252$). The shaded regions represent the 95% confidence intervals.

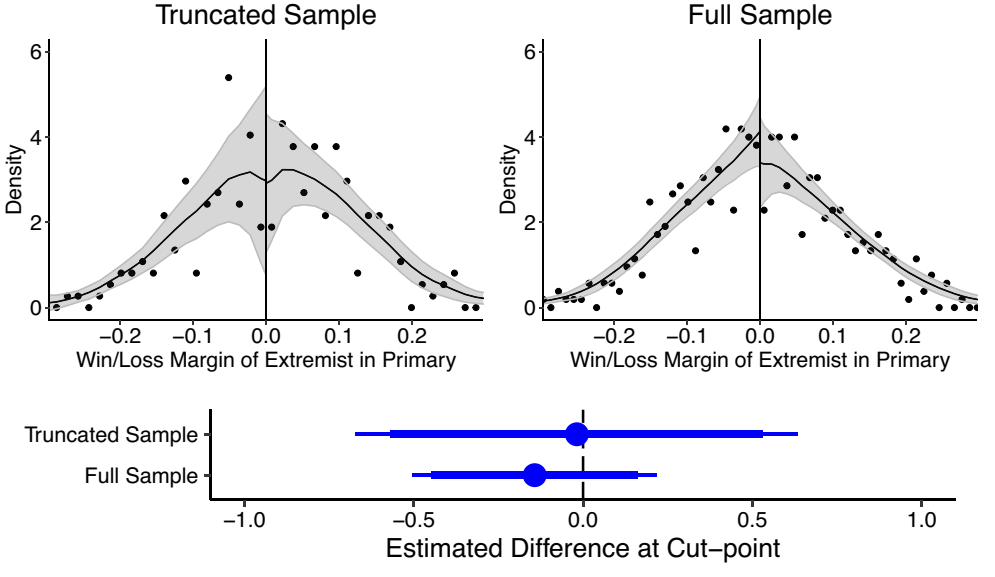

Figure 3. McCrary density test for elections between extremist and moderate candidates (1980–2010). The left-hand graph shows the results for the elections where the top-two candidates were at or above the median ideological distance ($n = 252$). The right-hand graph shows the results for the full sample ($n = 504$). The bottom coefficient plot shows the estimated differences at the cut-point in the above two graphs, along with the 95% and 90% confidence intervals for the estimated differences (thin and thick lines).

the narrower confidence intervals. In particular, the imbalance on Northeast found in the truncated sample is not evident in the full sample, suggesting that it may have just been a chance imbalance.

Balance also looks fairly good for our measure of predicted treatment probability. To construct this variable, we predicted an extremist primary victory using logistic regression on the data 5% or more from the cut-point. We then used this logistic regression model to estimate the likelihood of an extremist primary win for all elections in the dataset. The predictor variables that we used were those without

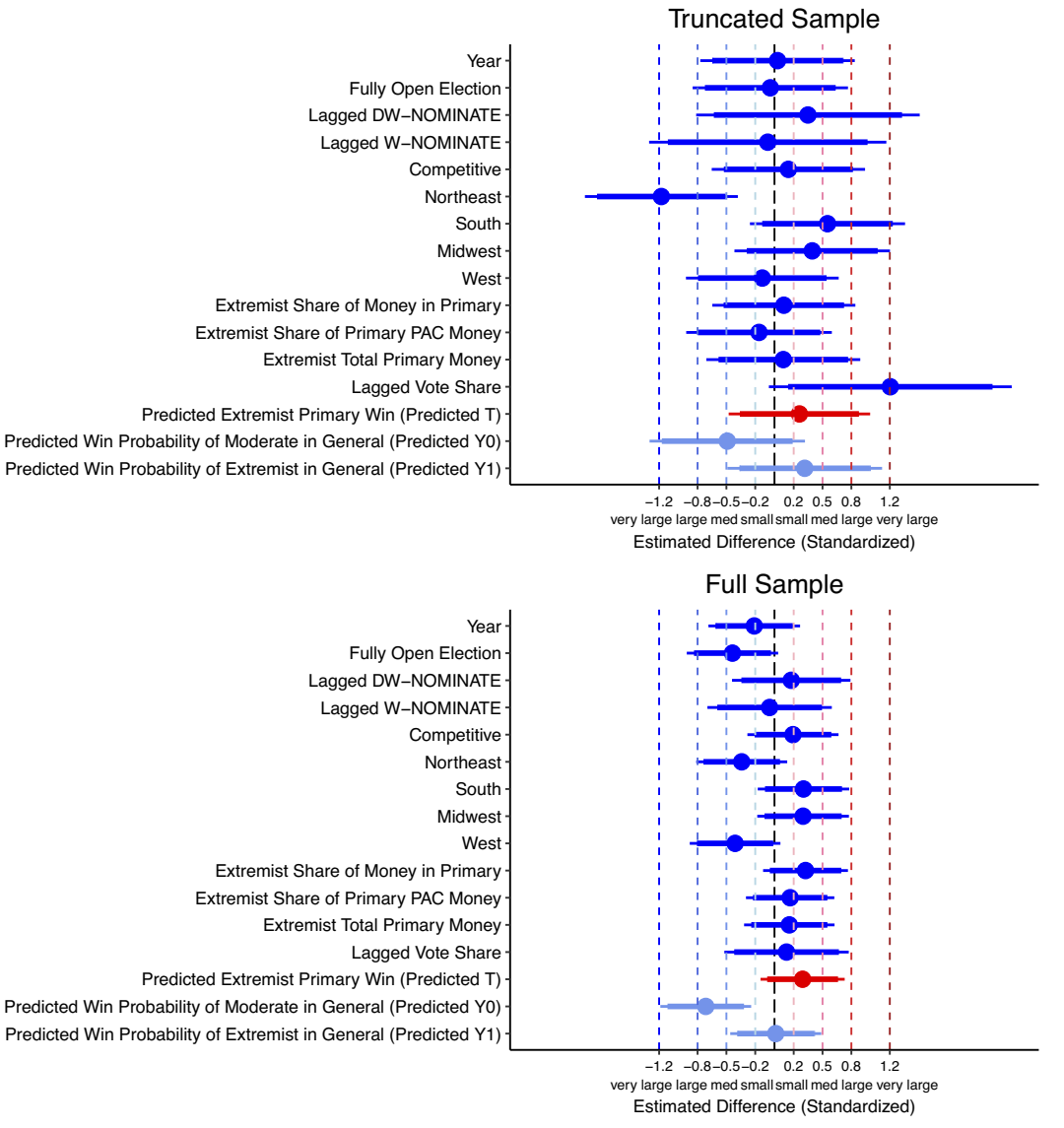

Figure 4. Illustrating balance between cases where extremist and moderate candidates barely won (1980–2010). The top coefficient plot shows the results for the elections where the top-two candidates were at or above the median ideological distance (n = 252), whereas the bottom coefficient plot shows the results for the full sample (n = 504). The thin lines represent the 95% confidence intervals, and the thick lines represent the 90% confidence intervals.

missingness: year, fully open general election, extremist share of money in the primary, and region. The point estimate and confidence interval (in red) indicate that there seems to be fairly good balance on this variable in the truncated data, though more potential concern in the full dataset. It is also notable that the estimates point in the more concerning direction. Specifically, in the cases where the "more extreme" candidate barely won, the pre-election covariates suggest that those were cases where the "more extreme" candidate had a higher chance of winning to begin with.

If this does reflect sorting, in what way might it bias the result? The light blue point estimates and confidence intervals in Figure 4 can give us some indication. These predicted potential outcomes come

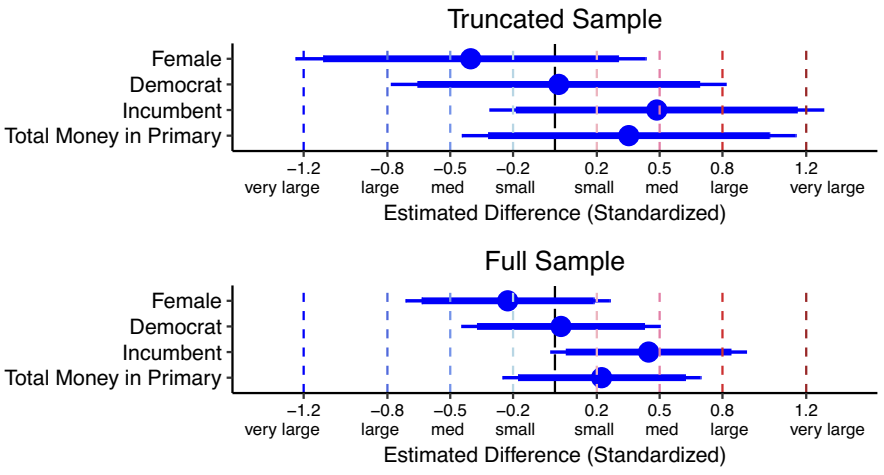

Figure 5. Exploring differences between extremist and moderate bare winners (1980–2010). The sample size in the top coefficient plot is 252 and in the bottom coefficient plot is 504. The thin lines represent the 95% confidence intervals, and the thick lines represent the 90% confidence intervals. See Hall (2015, 31) for a similar analysis, specifically on the truncated dataset.

from regression models on the data within 5% of the cut-point.[2] In the truncated data, there appears to be a jump at the cut-point in the predicted win probability of extremist candidates and a drop in the predicted win probability for moderate candidates, although neither difference is statistically significant. This suggests that more formidable extremist candidates may have been somewhat more likely to barely win, and less formidable moderate candidates may have been somewhat more likely to barely lose. In other words, stronger extremist and moderate candidates may have been more likely to win the close elections. However, the two sources of bias caused by this sorting point in opposite directions. Because the anticipated bias is an average of these two oppositely-signed imbalances (Equation 5), they would largely cancel each other out.

This example points to a strength of the PCRD: in a traditional RD setting, such as estimating how winning an election affected a candidate's wealth later in life, we would be very concerned if candidates with a larger share of the campaign money won at a much higher rate than candidates with a lower share of the campaign money. However, for a PCRD, such potential sorting might be less concerning, depending on how it relates to the treatment variable. If more well-funded candidates of both type A and type B were more likely to win close elections, then we should have some degree of balance on campaign money despite this sorting issue. Moreover, where we can estimate the possible bias because it is rooted in observables, we can get insights into how biased the overall PCRD estimate might be due to this specific concern.

As this approach does not isolate the effect of extremism as apart from other leader characteristics, additional analyses may be of interest to explore how extremism relates to other factors that might help to explain extremists' (reduced) success in general elections. Figure 5 compares the extremist primary winners to the moderate primary winners on four factors for which we have data. This analysis follows Hall (2015, 31). Extremists and moderates had a similar share of money in the primary, suggesting that the amount of money ultimately available is not likely a source of this difference (though certainly the strategies that the candidates decided to employ to obtain that funding could be important). The only

[2]For instance, to obtain the predicted win probabilities of the "more moderate" candidates in the general election (the $Y_i(0)$ potential outcome), we applied a logistic regression model to the data in the 45%–50% range (where the "more moderate" candidate won the primary) to determine which factors predicted a moderate victory in the general election. We then used this model to predict moderates' likelihood of winning the general election for all the cases, including those where the "more extremist" candidate won the primary election. Our predictor variables were again year, fully open general election, extremist share of money in the primary, and region.

notable potential difference was incumbency (which was not statistically significant at the 5% level). In the full sample, extremists were slightly more likely to be incumbents. This would only help explain the poorer performance of extremists in general elections if in the types of elections occurring in these districts, there was an incumbency disadvantage.

5. Conclusion

In this article, we emphasize and clarify that PCRDs estimate the LATEs for districts of electing certain types of politicians, not the effects of politician attributes. Further, we clarify why understanding the LATEs of electing certain types of politicians is valuable for both theoretical and policy reasons. We also address some points of confusion regarding PCRDs, in particular the tendency to think about them as only balancing district-level covariates. We then outline a number of tools that can be useful for researchers using PCRDs or gradient PCRDs. Finally, we apply these tools to the important empirical case of extremist candidates winning primaries for U.S. house races (Hall 2015).

Our article complements the recent work of Marshall (2024) in bringing attention to the PCRD, its distinction from other close-election research designs, and cause for skepticism or concern over what is identified by this design. In contrast to Marshall (2024), we see PCRDs as promising and valuable for answering the right types of questions, and we seek to offer researchers useful tools for analyzing and interpreting the results from such designs. In particular, we address several critiques that Marshall (2024) and others raise concerning PCRDs, such as the issue of compensating differentials and questions about the interpretability of bundled treatments and their usefulness for testing specific theories. We argue that when PCRDs are correctly interpreted, these concerns are either avoided, or are analogous to well-understood concerns present in even randomized experiments.

In recent years, the PCRD approach has provided many important contributions to political science, and it can continue to do so in the future. It should not be cast aside due to its inability to isolate the causal effect of the attribute of interest, which may be incoherent, undefined, or uninteresting in many contexts anyway. Nor should it be abandoned due to risks that candidates in close elections will differ in important ways from candidates in less-competitive races. Rather, these two issues constitute important considerations that should be carefully taken into account when interpreting the results from PCRDs.

Acknowledgments. We thank Robert Trager, John Marshall, Andrew Hall, Justin Grimmer, Yiqing Xu, Shiro Kuriwaki, Anton Strezhnev, Erin Hartman, Dan Thompson, Nina McMurry, David Ami Wulf, Alexander Kuo, Guillermo Toral, Michael Becher, Guadalupe Tuñón, Jasjeet Sekhon, Josh Kalla, and David Broockman for helpful comments and discussions. We also thank Jeff Gill and the two anonymous reviewers for valuable feedback.

Competing interests. The authors declare none.

Funding statement. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement. Replication code for this article is available at Bertoli and Hazlett (2025). A preservation copy of the same code and data can also be accessed via Dataverse at https://doi.org/10.7910/DVN/SCTVOF.

Supplementary Material. For supplementary material accompanying this paper, please visit http://doi.org/10.1017/pan.2025.5.

References

Alonso, J. M., and R. Andrews. 2020. "Political Ideology and Social Services Contracting: Evidence from a Regression Discontinuity Design." *Public Administration Review* 80 (5): 743–754.

Beland, L.-P. 2015. "Political Parties and Labor-Market Outcomes: Evidence from US States." *American Economic Journal: Applied Economics* 7 (4): 198–220.

Benedictis-Kessner, J. de, and C. Warshaw. 2016. "Mayoral Partisanship and Municipal Fiscal Policy." *The Journal of Politics* 78 (4): 1124–1138.

Benedictis-Kessner, J. de, and C. Warshaw. 2020. "Accountability for the Local Economy at all Levels of Government in United States Elections." *American Political Science Review* 114 (3): 660–676.

Bertoli, A., A. Dafoe, and R. F. Trager. 2019. "Is There a War Party? Party Change, the Left–Right Divide, and International Conflict." *Journal of Conflict Resolution* 63 (4): 950–975.
Bertoli, A., A. Dafoe, and R. F. Trager. 2024. "Leader Age and International Conflict: A Regression Discontinuity Analysis." *Journal of Peace Research* 61 (4): 643–658.
Bertoli, A., and C. Hazlett. 2025. "Replication Data for: Seeing Like a District: Understanding What Close-Election Designs for Leader Characteristics can and Cannot Tell us." Harvard Dataverse, V1. https://doi.org/10.7910/DVN/SCTVOF.
Bhalotra, S., and I. Clots-Figueras. 2014. "Health and the Political Agency of Women." *American Economic Journal: Economic Policy* 6 (2): 164–197.
Bhalotra, S., I. Clots-Figueras, G. Cassan, and L. Iyer. 2014. "Religion, Politician Identity and Development Outcomes: Evidence from India." *Journal of Economic Behavior & Organization* 104: 4–17.
Bhalotra, S., I. Clots-Figueras, and L. Iyer. 2018. "Pathbreakers? Women's Electoral Success and Future Political Participation." *The Economic Journal* 128 (613): 1844–1878.
Brollo, F., and T. Nannicini. 2012. "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil." *American Political Science Review* 106 (4): 742–761.
Brollo, F., and U. Troiano. 2016. "What Happens When a Woman Wins an Election? Evidence From Close Races in Brazil." *Journal of Development Economics* 122: 28–45.
Broockman, D., and J. Kalla. 2016. "Durably Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing." *Science* 352 (6282): 220–224.
Broockman, D. E. 2014. "Do Female Politicians Empower Women to Vote or Run for Office? A Regression Discontinuity Approach." *Electoral Studies* 34: 190–204.
Broockman, D. E., and T. J. Ryan. 2016. "Preaching to the Choir: Americans Prefer Communicating to Copartisan Elected Officials." *American Journal of Political Science* 60 (4): 1093–1107.
Callen, M., S. Gulzar, and A. Rezaee. 2020. "Can Political Alignment be Costly?" *The Journal of Politics* 82 (2): 612–626.
Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.
Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015. "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7 (1): 38.
Carozzi, F., and A. Gago. 2023. "Who Promotes Gender-Sensitive Policies?" *Journal of Economic Behavior & Organization* 206: 371–405.
Casarico, A., S. Lattanzio, and P. Profeta. 2022. "Women and Local Public Finance." *European Journal of Political Economy* 72: 102096.
Clots-Figueras, I. 2012. "Are Female Leaders Good for Education? Evidence from India." *American Economic Journal: Applied Economics* 4 (1): 212–244.
Dube, O., and S. Harish. 2020. "Queens." *Journal of Political Economy* 128 (7): 2579–2652.
Dynes, A. M., and J. B. Holbein. 2020. "Noisy Retrospection: The Effect of Party Control on Policy Outcomes." *American Political Science Review* 114 (1): 237–257.
Erikson, R. S., O. Folke, and J. M. Snyder Jr. 2015. "A Gubernatorial Helping Hand? How Governors Affect Presidential Elections." *The Journal of Politics* 77 (2): 491–504.
Feierherd, G. 2020. "How Mayors Hurt Their Presidential Ticket: Party Brands and Incumbency Spillovers in Brazil." *The Journal of Politics* 82 (1): 195–210.
Ferreira, F., and J. Gyourko. 2014. "Does Gender Matter for Political Leadership? The Case of US Mayors." *Journal of Public Economics* 112: 24–39.
Fowler, A., and A. B. Hall. 2017. "Long-Term Consequences of Election Results." *British Journal of Political Science* 47 (2): 351–372.
Fredriksson, P. G., L. Wang, and P. L. Warren. 2013. "Party Politics, Governors, and Economic Policy." *Southern Economic Journal* 80 (1): 106–126.
Gagnon-Bartsch, J., and Y. Shem-Tov. 2019. "The Classification Permutation Test." *The Annals of Applied Statistics* 13 (3): 1464–1483.
Gerber, E. R., and D. J. Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55 (2): 326–339.
Glynn, A. N., and M. Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science* 59 (1): 37–54.
Hall, A. B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109 (1): 18–42.
Hall, A. B. 2019. *Who Wants to Run? How the Devaluing of Political Office Drives Polarization*. Chicago: University of Chicago Press.
Hazlett, C. 2020. "Kernel Balancing." *Statistica Sinica* 30 (3): 1155–1189.
Healy, A., and N. Malhotra. 2013. "Childhood Socialization and Political Attitudes: Evidence from a Natural Experiment." *The Journal of Politics* 75 (4): 1023–1037.
Hopkins, D. J., and K. T. McCabe. 2012. "After its Too Late: Estimating the Policy Impacts of Black Mayoralties in US Cities." *American Politics Research* 40 (4): 665–700.

Huidobro, A., and A. Falcó-Gimeno. 2023. "Women Who Win But Do Not Rule: The Effect of Gender in the Formation of Governments." *The Journal of Politics* 85 (4): 1562–1568.
Imbens, G. W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–635.
Jankowski, M., K. Marcinkiewicz, and A. Gwiazda. 2019. "The Effect of Electing Women on Future Female Candidate Selection Patterns: Findings from a Regression Discontinuity Design.' *Politics & Gender* 15 (2): 182–210.
Lee, D. S., E. Moretti, and M. J. Butler. 2002. "Are Politicians Accountable to Voters? Evidence from US House Roll Call Voting Records." Working Paper.
Leigh, A. 2008. "Estimating the Impact of Gubernatorial Partisanship on Policy Settings and Economic Outcomes: A Regression Discontinuity Approach." *European Journal of Political Economy* 24 (1): 256–268.
Marshall, J. 2024. "Can Close Election Regression Discontinuity Designs Identify Effects of Winning Politician Characteristics?" *American Journal of Political Science* 68 (2): 494–510.
McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.
Mousa, S. 2020. "Building Social Cohesion Between Christians and Muslims Through Soccer in Post-ISIS Iraq." *Science* 369 (6505): 866–870.
Novaes, L. M. 2018. "Disloyal Brokers and Weak Parties." *American Journal of Political Science* 62 (1): 84–98.
Núñez, A. R., and E. Dinas. 2023. "Mean Streets: Power, Ideology and the Politics of Memory." *Political Geography* 103: 102840.
Patterson, S., Jr. 2020. "Estimating the Unintended Participation Penalty Under Top-Two Primaries with a Discontinuity Design." *Electoral Studies* 68: 102231.
Sekhon, J. S., and R. Titiunik. 2012. "When Natural Experiments are Neither Natural Nor Experiments." *American Political Science Review* 106 (1): 35–57.
Szakonyi, D. 2021. "Private Sector Policy Making: Business Background and Politicians Behavior in Office." *The Journal of Politics* 83 (1): 260–276.
Thomas, A. 2018. "Targeting Ordinary Voters or Political Elites? Why Pork is Distributed Along Partisan Lines in India." *American Journal of Political Science* 62 (4): 796–812.
Thompson, D. M. 2020. "How Partisan is Local Law Enforcement? Evidence from Sheriff Cooperation with Immigration Authorities." *American Political Science Review* 114 (1): 222–236.
VanderWeele, T. J., and M. A. Hernan. 2013. "Causal Inference Under Multiple Versions of Treatment." *Journal of Causal Inference* 1 (1): 1–20.
Wainstein, L. 2022. "Targeted Function Balancing." Preprint, arXiv:2203.12179.
Waltz, K. N. 2010. *Theory of International Politics*. Long Grove: Waveland Press.
Xu, Y., and Y. Yao. 2015. "Informal Institutions, Collective Action, and Public Investment in Rural China." *American Political Science Review* 109 (2): 371–391.

Cite this article: Bertoli, A. and Hazlett, C. (2025). Seeing Like a District: Understanding What Close-Election Designs for Leader Characteristics Can and Cannot Tell Us. *Political Analysis*, 393–411. https://doi.org/10.1017/pan.2025.5